大豆 *GeBP* 转录因子基因家族的生物信息学分析

巩元勇*，赵丽华， 闫 飞，朱丽红

(攀枝花学院 生物与化学工程学院，四川 攀枝花 617000)

摘要：GeBP 转录因子调控植物表皮毛的生长发育，并且参与控制植物叶片的发育。利用生物信息学方法，在大豆全基因组范围内搜索 *GeBP* 基因家族，并从氨基酸理化性质、基因结构、染色体的物理分布、系统进化、序列比对、功能结构域、组织表达情况等基本特征方面对 *GmGeBP* 基因家族进行分析。共获得 9 个 *GmGeBP* 转录因子家族基因成员，其中仅 2 个基因含有内含子，而且都只有 1 个内含子，表明该家族成员基因构造比较简单但稳定。*GmGeBP* 编码的蛋白分子量为 39.65～49.24 kD，理论等电点为 4.65～9.08；这些成员基本上都是酸性氨基酸，属于亲水性、不稳定蛋白。这 9 个基因不均匀的分布于 7 条染色体上，10 和 20 号染色体上分别分布 2 个 *GeBP* 基因，3、5、13、15、19 号染色体上各分布 1 个基因。系统进化分析表明，大豆与拟南芥对应的 *GeBP* 成员亲缘关系较近，分别聚类到 4 个分支，而与水稻的距离较远。结构域分析表明，9 个 GmGeBP 成员都包含 DUF573 结构域，推测该部分在 GeBP 转录因子中很可能是与靶标基因顺式作用元件互作的结构域。通过分析大豆 *GmGeBP* 转录因子家族基因的组织表达，发现不同基因的在大豆不同组织的表达量不同，具有一定的特异性。该文对大豆 *GeBP* 转录因子基因家族的分析和鉴定为进一步研究大豆表皮毛发育的分子作用提供了理论基础。

关键词：大豆，*GeBP* 转录因子，生物信息，组织表达

中图分类号： **文献标识码：**A 文章编号：

Bioinformatics analysis of *GeBP* transcription factor gene family in soybean

GONG Yuanyong*, ZHAO Lihua, YAN Fei, ZHU Lihong

(*Biological and Chemical Engineering College, Panzhihua University*, Panzhihua 617000, Sichuan, China)

Abstract: It has been clarified that GeBP transcription factor regulates the growth and development of plant epidermal hair and participates in the control of plant leaf development. The bioinformatics methods were used to identify the *GeBP* gene family in the whole soybean genome, and from physicochemical properties of amino acids, as well as gene structure, physical distribution of chromosomes, phylogenetic tree, and multiple sequence comparison, the functional domain, tissues expression and other basic characteristics of *GmGeBP* gene family were analyzed.

收稿日期：

基金项目：国家自然科学基金项目(31301682)；金沙江干热河谷生态修复与治理创新研究团队专项经费(035200179)；攀枝花大学科技园发展有限责任公司种子资金"双创"项目(2019-46) [Supported by the National Natural Science Foundation of China (31301682); Special Fund for Research Team of Ecological Restoration and Governance Innovation in Dry Hot Valley of Jinsha River (035200179); "Double Creation" Project of Seed Fund of Panzhihua University Science Park Development Co., Ltd. (2019-46)]

作者简介：巩元勇(1982 -)，男，山东省惠民县，博士，副教授，主要从事植物生物技术研究，（E-mail）gyy2011qh@163.com。

*通信作者

A total of nine members of *GmGeBP* transcription factor family were identified, of which only two genes contained introns and all had only one intron, indicating that the gene structure of the family members was relatively simple but stable. The molecular weight of GmGeBPs is 39.65-49.24 kD, and the theoretical isoelectric point is 4.65-9.08; these members are basically acidic amino acids, which are hydrophilic and unstable proteins.The chromosome physical distribution showed that 9 genes were unevenly distributed on 7 chromosomes, 2 *GeBP* genes on chromosome 10 and 20，respectively, and 1 gene on chromosome 3, 5, 13, 15 and 19 respectively. The phylogenetic analysis showed that GeBP members of soybean and *Arabidopsis thaliana* were closely related, clustered into four branches respectively, but far away from *Oryza sativa*. The analysis of domains showed that all the nine GmGeBP members contained DUF573 domain, which was probably the domain interacting with cis-acting elements of target genes in GeBP transcription factors. By analyzing the expression of *GmGeBP* transcription factor family, we found that the expression of different genes in different tissues is different, with a certain specificity. The analysis and identification of *GmGeBP* transcription factor gene family provided a theoretical basis for further studying the molecular role of soybean epidermal development.
Key words: *Glycin max*, *GeBP* transcription factor, bioinformatics, tissue expression

表皮毛广泛分布于陆地植物的叶片、茎秆以及花萼等地上部器官的表面，是植物表皮细胞分化形成的一种特殊的细胞形态。表皮毛是植物的第一道保护屏障，通过调节水分的蒸腾作用，减缓叶片的热负荷，增强对冷冻或紫外线的耐受性，增强植物抵御昆虫捕食的防御能力。转录因子是生物体生长发育过程中一类重要的调节因子，高等生物体因其机体的复杂性需要更多的转录因子参与。GeBP(GLABROUS1 enhancer binding protein)是一类植物特有的转录因子，拟南芥 GeBP 可以通过与 *GL1*(*GLABROUS1*)基因的互作调控来控制表皮毛的发生(普利等，2003)。GeBP 包含中央 DNA 结合区、bZIP 转录因子保守域和 C 末端保守区，然而这个 bZIP 转录因子保守域有别于经典的 bZIP 转录因子保守域，且真正发挥功能的是中央 DNA 结合区和 C 末端保守区，所以 GeBP 是植物中一类新的转录因子蛋白(Curaba et al., 2003；Chevalier et al., 2008)。当前只在拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa*)、番茄(*Solanum lycopersicum*)和毛竹(*Phyllostachys edulis*)中有关于 *GeBP* 转录因子基因家族的报道，这四种植物 *GeBP* 基因家族成员分别是 22 个(Chevalier et al., 2008)、15 个(石蕾，2013)、10 个(陈凯等，2019)和 16 个(单雪萌等，2020)，关于 GeBP 转录因子具体功能的报道还很少。

在植物中，转录因子通常参与激素途径与激素互作来调控植物的发育。GeBP 蛋白可以与 *GL1* 基因顺式调控元件结合调控该基因的转录，*GL1* 基因属于 *myb* 基因，参与表皮细胞决定且被赤霉素和细胞分裂素调控(Gan et al., 2007)；*GeBP* 基因的表达受到 *KNOX* 家族转录因子 *BP*(*BREVIPEDICELLUS*)基因的正向调控(Curaba et al., 2004)，KNOX 在茎端分生组织正向调控细胞分裂素途径(Jasinski et al., 2005)，由此推测 GeBP 可能通过调控赤霉素和细胞分裂素途径来控制表皮毛的发生(Chevalier et al., 2008)。

Ray et al.(2011)研究发现 C_2H_2、C_2C_2、C_3H、*LIM*、*PHD*、*WRKY*、*ZF-HD* 和 *ZIM* 等锌指类转录因子成员，以及 *GeBP*、*jumonji*、*MBF1* 和 *ULT* 等转录因子家族在缺水条件下出现表达差异。MBF1，jumonji，ULT 和 GeBP 这四类转录因子家族一般认为主要参与植物发育过程和植物激素反应(Curaba et al., 2003；Noh et al., 2004；Kenichi et al., 2004；Carles et al., 2005；Chevalier et al., 2008)，它们通常不参与胁迫条件下的应激反应性，然而在水分亏缺胁迫条件下都表现为表达上调，表明 GeBP 等转录因子在植物对干旱逆境的反应中发挥一定的作用。新近研究发现，LiGeBP 和 LiMYB、LibZIP、LieBp-2、LiERF 等五类转录因子可能是薰衣草中单萜合酶的激活剂，LiGeBP 可能参与控制薰衣草中单萜合酶的合成(Sarker et al., 2019)。

大豆(*Glycin max*)的起源在中国，随着人类活动范围的扩大已经广泛种植于世界各地，

大豆早就成为世界性的重要经济作物，它为我们提供了主要的植物油和植物蛋白，同时也是动物饲料蛋白质的主要来源。大豆表皮毛同其它植物的一样也是典型的单细胞结构，不存在分支。研究证实，大豆表皮毛的密度同抗虫和抗旱等性状关系密切，但是关于大豆表皮毛发育的分子基础研究还鲜有报道。大豆基因组测序工作已于 2010 年完成并公布(Schmutz et al., 2010)，这为大豆相关基因家族及功能基因在基因组水平上的探索研究提供了可能。大豆 *WRKY*、*ERF*、*Dof* 等转录因子基因家族在全基因组层面的分析报道也越来越多(Yu et al., 2016；Song et al., 2016；翟莹等，2016；翟莹等，2019；刘蓓等，2020)，但是还未见有关于大豆 *GeBP* 转录因子基因家族的研究报道。本研究通过生物信息学的方法，从基因的核苷酸序列长度和氨基酸序列的基本理化性质、基因在染色体的物理定位、基因结构、系统进化树、序列对比、功能结构域分布、基因在不同组织的表达情况等基本特征方面对大豆 *GeBP* 转录因子家族进行全面预测和分析，为进一步深入探究大豆 *GeBP* 转录因子家族基因的生理生化功能提供理论依据。

1 材料与方法

1.1 材料

从植物转录因子数据库 PlantTFDB(http://planttfdb.cbi.pku.edu.cn/)搜索获得拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa*)和大豆(*Glycine max*)这 3 个物种 *GeBP* 转录因子基因家族共 40 个成员的基因座位置信息，然后从 JGI 的 Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html)搜索获得水稻、大豆和拟南芥的 *GeBP* 转录因子基因的编码区及 CDS 序列和氨基酸序列，同时获得大豆 *GeBP* 转录因子基因家族在不同组织部位表达的 FPKM(Fragments Per Kilobase of transcript per Million fragments mapped)值。

1.2 方法

1.2.1 大豆 *GeBP* 基因家族基本信息获取

利用 DNASTAR Lasergene 软件中的 EditSeq 分析所获得大豆 *GeBP* 基因的编码区和 CDS 序列长度，用 Expasy（https://web.expasy.org/protparam/）在线分析大豆 GeBP 的氨基酸序列，获得氨基酸残基长度、分子质量、理论等电点、不稳定系数、亲水性指数等基因基本信息。

1.2.2 大豆 *GeBP* 基因染色体定位

大豆 *GeBP* 基因在染色体的位置信息来自 Phytozome 大豆基因组数据库，从 NCBI 的 Genome Data Viewer（https://www.ncbi.nlm.nih.gov/genome/gdv/）获得大豆每条染色体的总长度，根据这些信息用 maplnspect 软件绘制大豆 *GeBP* 基因在染色体上的物理分布图。

1.2.3 大豆 *GeBP* 基因的生物信息学分析

用 GSDS9(Gene Structure Dispely Server，http://gsds.cbi.pku.edu.cn/)(Hu et al., 2015)在线绘制大豆和拟南芥 *GeBP* 基因结构图；用 MEGA7 软件采用邻接法 NJ(Neighbor-Joining)构建大豆、拟南芥和水稻 *GeBP* 基因家族氨基酸序列的系统进化树(Kumar et al., 2018)，校验参数 Bootstrap=1000；运用 DNAMAN 软件对大豆 *GeBP* 基因家族的氨基酸序列进行多重序列比对；用 Lasergene 软件的 MegAlign 分析大豆 *GeBP* 基因家族氨基酸序列间的相似性；用 Pfam 32.0(http://pfam.xfam.org/)在线搜寻鉴定大豆 GeBP 蛋白序列功能结构域(Domain)的存在情况。

1.2.4 大豆 *GeBP* 基因在不同组织的表达

用 Heml 软件(http://hemi.biocuckoo.org/down.php)绘制大豆 *GeBP* 基因在花(Flower)、叶(Leaves)、根瘤(Nodules)、荚果(Pod)、根(Root)、根毛(Root Hair)、种子(Seed)、茎尖分生组织(Shoot Apical Meristem)、茎(Stem)等 9 个不同组织的表达热图。

2 结果与分析

2.1 大豆 *GeBP* 基因的基本信息及染色体定位

通过在植物转录因子数据库 PlantTFDB 搜索获得大豆转录因子 GeBP 基因成员，共获得

9 个大豆无表皮毛增强子结合蛋白基因，分别命名为 *GmGeBP1-9*(表 1)。 在 Phytozome 大豆基因组数据库搜索获得这 9 个基因所对应的编码区序列、CDS 序列、氨基酸序列，并用 Expasy 在线分析氨基酸序列获得蛋白质序列基本的理化性质信息。

表 1 大豆 *GeBP* 基因家族成员基本信息

Tab.1 Basic information of *GeBP* gene family in *Glycine max*

基因名称 Gene name	基因座 ID Locus ID	编码区长度 Length of coding region /bp	CDS 长度 Length of CDS /bp	蛋白质长度 Protein Length/aa	分子质量 MW/kD	理论等电点 Isoelectric point/pI	不稳定系数 Instability Index	亲水性指数 Hydropathy Index
GmGeBP1	Glyma.03G245200	1218	1218	40	44.61	5.00	60.84	-0.752
GmGeBP2	Glyma.05G088300	1294	1164	387	43.23	9.08	46.87	-0.950
GmGeBP3	Glyma.10G160000	1119	1119	372	41.29	4.75	53.99	-0.706
GmGeBP4	Glyma.10G160100	1160	1128	375	43.15	5.20	55.57	-0.749
GmGeBP5	Glyma.13G251000	1347	1347	448	49.24	5.26	65.50	-1.221
GmGeBP6	Glyma.15G063300	1314	1314	437	48.35	5.48	69.85	-1.300
GmGeBP7	Glyma.19G242600	1215	1215	404	44.74	4.81	56.93	-0.805
GmGeBP8	Glyma.20G228300	1074	1074	357	41.06	4.93	63.02	-0.786
GmGeBP9	Glyma.20G228500	1062	1062	353	39.65	4.65	67.12	-0.778

如表 1 所示，*GmGeBP* 转录因子家族的 9 个成员分别分布在 7 条染色体上，其中 10 号染色体上有 *GmGeBP3* 和 *GmGeBP4* 两个 *GeBP* 基因，20 号染色体上有 *GmGeBP8* 和 *GmGeBP9* 两个 *GeBP* 基因。其余染色体各有一个 *GeBP* 基因。有 7 个大豆 *GeBP* 基因的编码区长度和 CDS 长度一样，表明这些 *GeBP* 基因结构中不含有内含子；只有 *GmGeBP2* 和 *GmGeBP4* 这两个基因的编码区长度和 CDS 长度不一样，说明这两个基因在基因结构上是包含内含子的。大豆 *GeBP* 基因翻译后蛋白质的长度在 353～448 个氨基酸之间，其中 5 个成员蛋白质长度在 300～400 个氨基酸之间；有 4 个家族成员的长度在 400～500 个氨基酸之间。其中，GmGeBP5 的氨基酸长度最长（448 aa），其分子质量也是最大，为 49.24 kD；GmGeBP9 的氨基酸长度最短（353 aa），其分子质量也是最小，为 39.65 kD。这 9 个 GeBP 蛋白的理论等电点也存在一定的差异，其理论等电点值在 4.65～9.08 之间变化，只有 GmGeBP2 的理论等电点为 9.08，是碱性氨基酸，其余 8 个 GeBP 蛋白理论等电点均小于 7，为酸性氨基酸。蛋白质的不稳定系数是用来分析该蛋白是否是稳定蛋白的，如果不稳定系数大于 40，则表明是不稳定蛋白，反之，如果不稳定系数小于 40，则表明是稳定蛋白，GmGeBP 所有的不稳定系数都大于 40，说明这些蛋白都是不稳定蛋白。亲水性指数大于+0.5 的为疏水性蛋白，亲水性指数若小于-0.5 的则为亲水性蛋白，如果介于－0.5～＋0.5 之间则为两性蛋白，所有 GmGeBP 蛋白的亲水性指数都小于-0.5，说明这些蛋白都是亲水性蛋白。

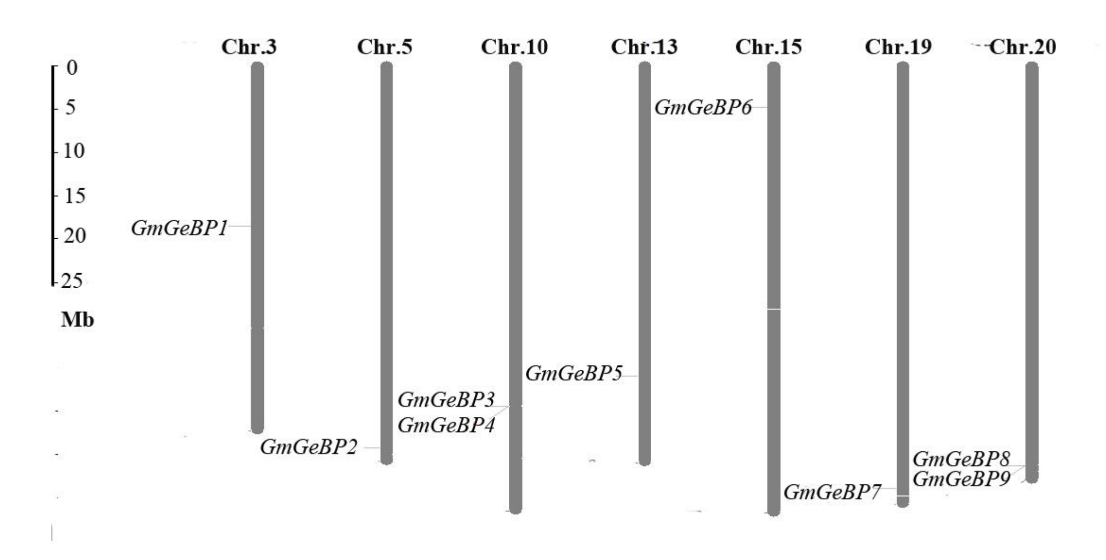


图 1 *GmGeBP* 基因的染色体物理定位

Fig.1 Chromosome physical location of *GmGeBP* genes

大豆共有 20 对 40 条染色体，基因在染色体上的分布用 1 号-20 号染色体来表示。大豆 *GeBP* 基因家族共有 9 个成员，不均匀的分布在 Chr3、Chr5、Chr10、Chr13、Chr15、Chr19、Chr20 等 7 条染色体上。其中在 Chr10 和 Chr20 上各分布有两个 *GeBP* 基因，从后面的序列一致性及进化树分析的结果来看，尽管 *GmGeBP3* 和 *GmGeBP4*、*GmGeBP8* 和 *GmGeBP9* 这两组基因在物理位置上距离很近，但是它们之间在进化上不存在复制关系，是独立进化的。在图 1 可以发现，这些基因在染色体上相对独立存在，在染色体上没有以基因簇的存在形式。

2.2 *GeBP* 基因结构分析

利用大豆和拟南芥 *GeBP* 基因家族成员的编码区序列和 CDS 序列通过 GSDS 在线构建基因结构图，用图像直观的来研究 *GeBP* 基因的结构情况。如图 2 所示，大豆 9 条 *GeBP* 基因有 7 个不包含内含子，另外两个有内含子的基因也分别包含一个内含子。在进行基因克隆研究基因功能的时候，没有内含子的基因可以直接通过提取的 DNA 为模板来获得基因序列，而没有必要提取 RNA 之后再反转录作为扩增模板。

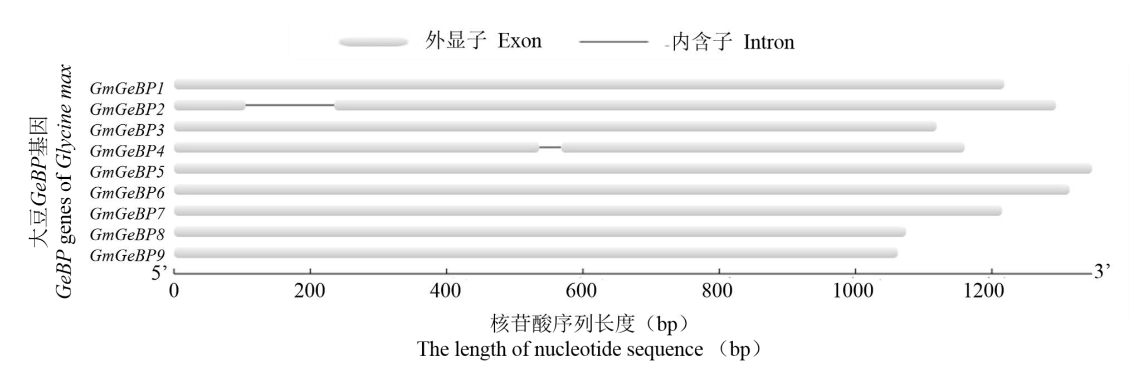


图 2 大豆 *GeBP* 基因结构图

Fig.2 Gene structure map of soybean *GeBP* genes

再来分析拟南芥 *GeBP* 基因家族基因结构，研究发现拟南芥 22 个 *GeBP* 基因有 16 个成员也没有内含子，有 6 个基因包含内含子，其中有 4 个基因均只含有 1 个内含子，1 个基因含有 3 个内含子，1 个基因含有 4 个内含子(图 3)。总体来看，*GeBP* 基因的基因结构在不同植物上都表现的比较稳定，因为没有内含子存在或存在很少的内含子，在转录时不易形成可变剪接体。

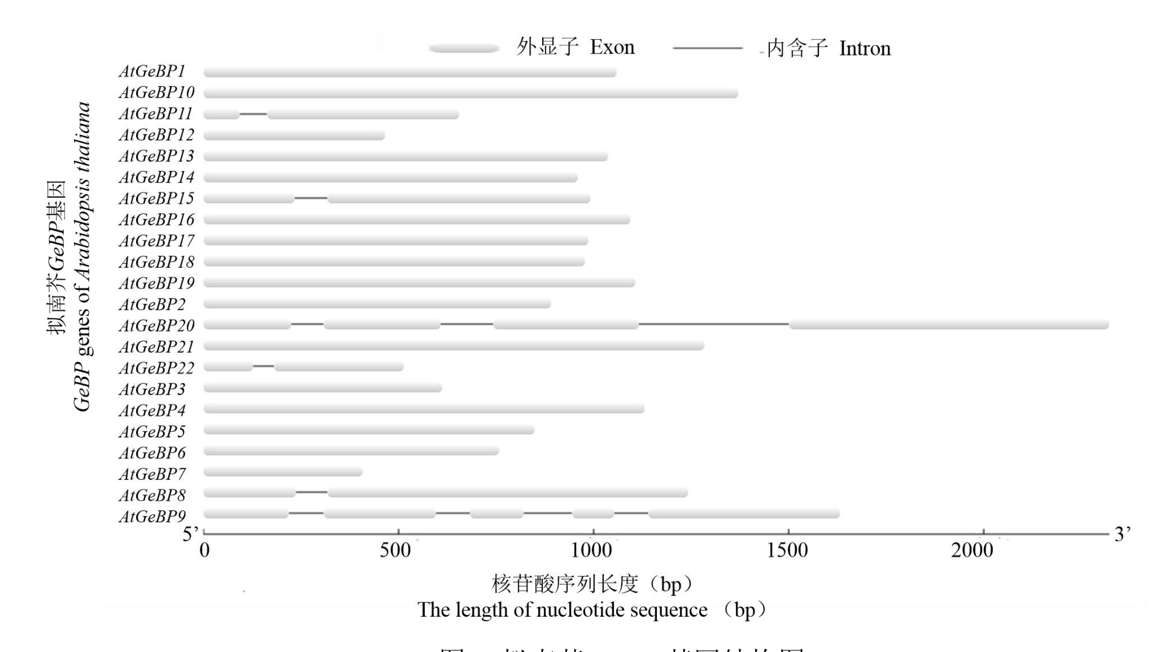


图 3 拟南芥 *GeBP* 基因结构图

Fig.3 GeBP genes structure map of *Arabidopsis thaliana*

2.3 构建 *GeBP* 基因家族进化树

为了进一步了解 *GeBP* 基因家族系统进化关系，本研究选取 20 个拟南芥 *GeBP* 基因、11 个水稻 *GeBP* 基因、9 个大豆 *GeBP* 基因等 3 个植物的 *GeBP* 基因进行比对分析，利用 MEGA 7 的邻接法构建这 40 个 GeBP 蛋白序列的进化树。其中，为了更加方便辨识，在图形处理上，将大豆基因标注黑色圆形，拟南芥基因标注白色圆形，水稻基因标注白色方框。如图 4 所示，进化树可以分为 4 个大的分支，分别含有 14 个、10 个、6 个和 10 个 *GeBP* 基因。在第一个分支，三个物种的基因都有；第二个分支只包含拟南芥的基因；第三个分支基因数量最少，但是包含有拟南芥和大豆的基因；第四个分支以水稻基因为主，仅有一个来源拟南芥的基因。可见，因大豆和拟南芥同属双子叶植物的缘故，这两个物种的基因在进化关系上要更近一些。

从单个基因的进化关系来看，相同物种间的基因进化关系最近，如大豆的 *GmGeBP1* 和 *GmGeBP7*，*GmGeBP3* 和 *GmGeBP9*，*GmGeBP4* 和 *GmGeBP8*，拟南芥的 *AtGeBP10* 和 *AtGeBP21*，*AtGeBP3* 和 *AtGeBP6*，*AtGeBP8* 和 *AtGeBP15*，水稻的 *OsGeBP6* 和 *OsGeBP9*，*OsGeBP3* 和 *OsGeBP5*，*OsGeBP4* 和 *OsGeBP8* 等。表明 *GeBP* 基因在物种间进化上比较保守。

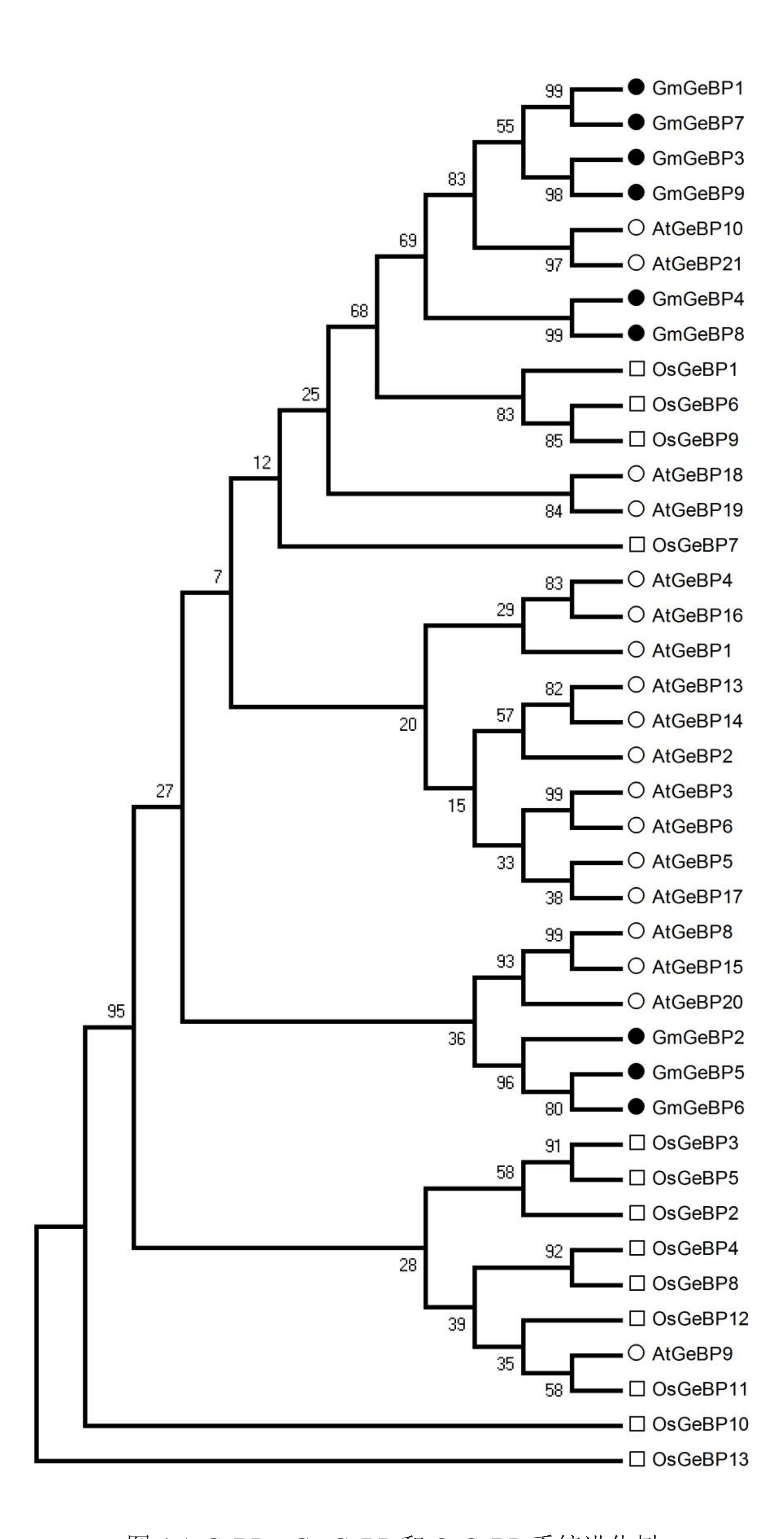


图 4 AtGeBP、GmGeBP 和 OsGeBP 系统进化树

Fig.4　The phylogenetic tree of AtGeBP、GmGeBP and OsGeBP proteins

2.4 大豆 *GeBP* 基因家族的序列比对和结构域分析

进行序列比对的目的是从核酸以及氨基酸的层次来分析序列的相同点和不同点，进而推测它们的结构、功能以及进化上的联系。大豆 GeBP 家族成员含有大约 131 个氨基酸组成的功能结构域，功能结构域含有大量的碱性氨基酸：精氨酸(R)和赖氨酸(K)。不同物种间 GeBP 蛋白的氨基酸序列的一致性比较低(结果没有展示)。大豆的 GeBP 蛋白质氨基酸多重

序列比对的结果表明，大豆本身的 GeBP 蛋白质氨基酸序列的一致性也不是很高，所有成员共有的氨基酸序列非常少(图 5)。

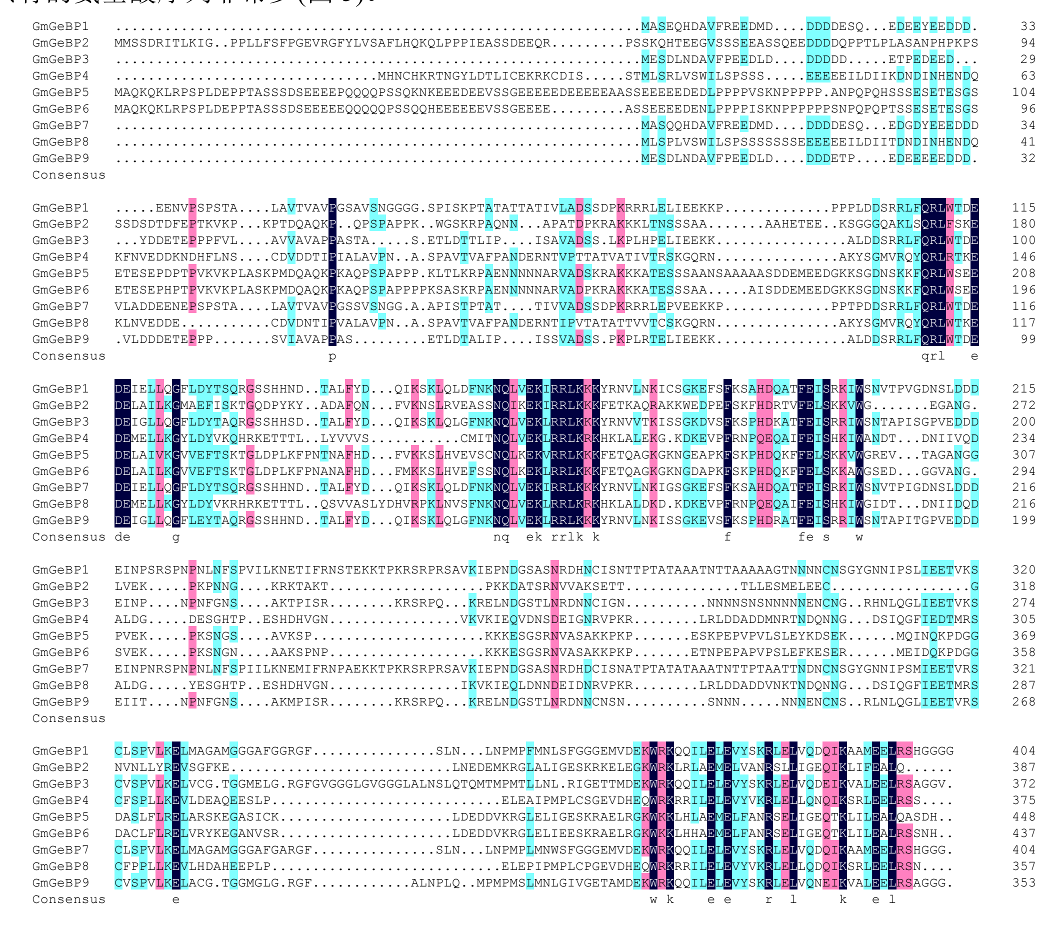


图 5 大豆 GeBP 蛋白质多序列对比图

Fig.5 Multi-sequence comparison diagram of soybean GeBP proteins

通过分析大豆 GeBP 基因家族成员之间的氨基酸序列的一致性发现，尽管整体序列一致性比较低，但是成员间的一致性还是有比较高的，*GmGeBP1* 和 *GmGeBP7* 的一致性最高，达到了 90.2%，其次是 *GmGeBP3* 和 *GmGeBP9* 的 88.5%，接着是 *GmGeBP4* 和 *GmGeBP8* 的 87.4%，还有 *GmGeBP5* 和 *GmGeBP6* 的 86.6%，这些都是一致性在 85%以上的序列，同进化树的分析结果也相似。氨基酸序列一致性高，表明在功能上这些基因也可能类似或互补。

如图 6 所示，大豆所有的 9 个 GmGeBP 成员都包含 DUF573 结构域，属于 DUF573 超家族。尽管标注为未知功能结构域，但是可以推测该部分在 *GeBP* 转录因子氨基酸序列中很可能与靶标基因顺式作用元件互作的结构域，但是不同 GmGeBP 中 DUF573 功能结构域所处的位置存在一定的差异，这可能是导致不同 GmGeBP 功能差异的原因之一。

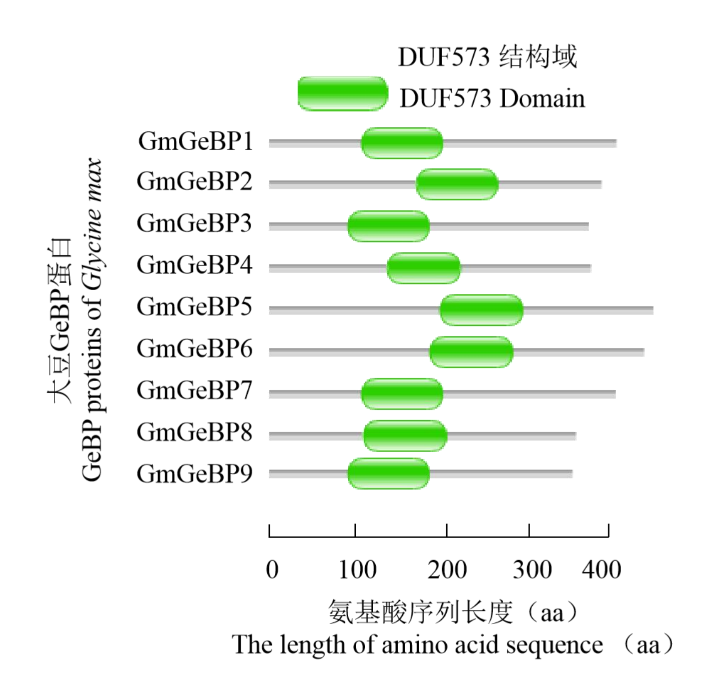
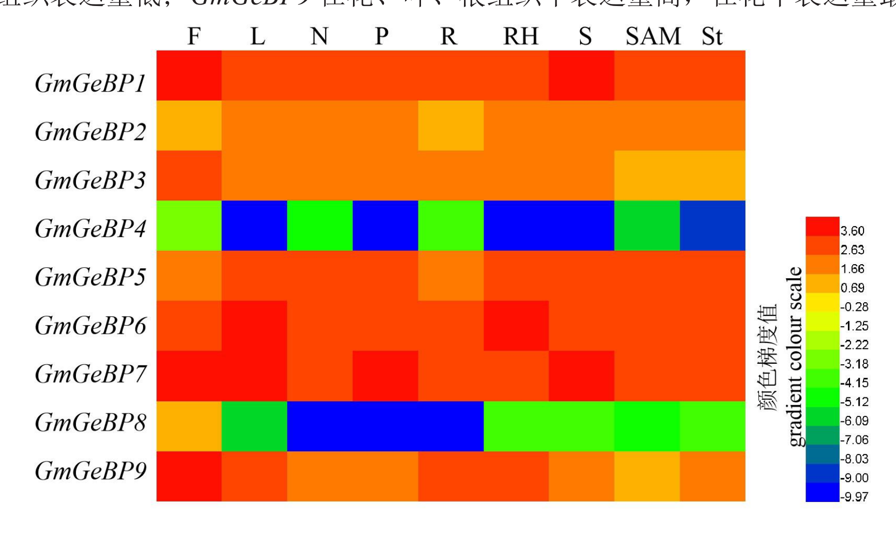


图 6 大豆 GeBP 蛋白功能结构域分布

Fig.6 Distribution of functional domains of GmGeBP proteins

2.5 大豆 *GeBP* 基因在不同组织的表达分析

总体来看，大豆 *GmGeBP* 家族成员在所有组织中都有表达，其中 *GmGeBP4* 和 *GmGeBP8* 在各个组织中的表达量相对其他基因都普遍偏低(图 7)。单个基因的表达情况来看，*GmGeBP1* 在所有组织中都表达，在花和种子中表达量最高，在叶、结节、荚果、根、根毛、茎尖分生组织、茎表达量次之；*GmGeBP2* 在叶、结节、荚果、根毛、种子、茎尖分生组织、茎的表达比在花和根中表达量更高；*GmGeBP3* 的表达主要集中在花器官中，在叶、结节、荚果、根、根毛、种子的表达中次之，在茎尖分生组织、茎的表达量更少；*GmGeBP4* 在各组织中表达量较低；*GmGeBP5* 在花、叶、结节、荚果、根、根毛、种子、茎尖分生组织、茎中均表达；*GmGeBP6* 在花、叶、结节、荚果、根、根毛、种子、茎尖分生组织、茎中均表达，在叶和根毛表达量最高；*GmGeBP7* 在花、叶、结节、荚果、根、根毛、种子、茎尖分生组织、茎中均表达，在花、叶、荚果、种子中表达最高；*GmGeBP8* 在花中表达量最高，在其他组织表达量低；*GmGeBP9* 在花、叶、根组织中表达量高，在花中表达量最高。



F. 花；**L.** 叶；**N.** 根瘤；**P.** 豆荚；**R.** 根；**RH.** 根毛；**S.** 种子；**SAM.** 顶端生长点；**St.** 茎

F. Flower; **L.** Leaves; **N.** Nodules; **P.** Pod; **R.** Root; **RH.** Root Hair; **S.** Seed; **SAM.** Shoot Apical Meristem; **St.** Stem

图 7 *GmGeBP* 基因在不同组织的表达热图

Fig.7 Expression heat map of *GmGeBP* genes in different tissue s

数据分析结果表明，同一基因在不同位置的表达量不一样，同一位置表达所对应的基因表达量也有差异，对此分析得到 *GmGeBP* 家族成员的表达具有一定的特异性，且各成员之间不尽相同。因此，一定程度上基因的功能取决基因在不同时期不同组织和器官的表达情况。

3 讨论与结论

对植物转录因子的研究，似乎更偏向于参与逆境胁迫反应途径相关的转录因子。以 *WRKY* 转录因子为例，*WRKY* 转录因子是植物体特有一类成员数量庞大的的转录因子家族，广泛的参与到植物对多种生物和非生物胁迫的反应过程(Jiang et al., 2017)，是当前研究最热的植物转录因子之一。2020 年 7 月 20 日在中国知网(https://www.cnki.net/)搜索篇名含有 WRKY 的文章，共找到 1512 条结果，不仅有数量众多的对不同植物种类 *WRKY* 转录因子基因家族的全基因组鉴定和分析，还有很多对单个 *WRKY* 基因相关功能的研究。反观对 GeBP 的搜索，只找到 7 篇文章，植物只涉及拟南芥、水稻、番茄和毛竹等 4 种，尽管最早关于 *GeBP* 基因研究的文章发表于 2003 年(Curaba et al., 2003)，但是这十几年来对 *GeBP* 转录因子的研究还是非常缓慢，*GeBP* 基因的很多功能还不明确。

从已经报道的结果来看，植物 *GeBP* 转录因子基因家族成员数量不是很多，拟南芥包含的 *GeBP* 成员最多，有 22 个(Chevalier et al., 2008)；毛竹含有 16 个成员(单雪萌等，2020)，水稻含有 15 个成员(石蕾，2013)，番茄含有 10 个成员(陈凯等，2019)；本文鉴定的大豆的 *GeBP* 成员更少，只有 9 个(表 1)。GeBP 转录因子家族成员基因结构相对简单，不含有内含子的基因所占比例很高。毛竹有 13 个成员没有内含子，这 3 个有内含子的基因，有 2 个只有 1 个内含子(单雪萌等，2020)；番茄有 8 个成员没有内含子，2 个有内含子的基因有 1 个只有 1 个内含子(单雪萌等，2020)；拟南芥有 16 个成员没有内含子，6 个有内含子的基因有 4 个只有 1 个内含子(图 3)；大豆有 7 个成员没有内含子，剩下有内含子的 2 个基因都只有 1 个内含子(图 2)。可见 *GeBP* 转录因子家族基因由于多数成员不含内含子，在转录时减少了出现可变剪接体的几率，基因在进化时就更加的稳定和保守。

在进化关系上，本文的研究结果同其它研究类似，单-双子叶植物之间各自的进化关系最近，同时都有一个分支为双子叶植物所特有，说到最末端分支，相同物种间的家族成员的进化关系要高于不同物种间的进化关系(陈凯等，2019；单雪萌等，2020；图 4)，这些都表明 *GeBP* 转录因子家族基因在进化上的保守性。所有报道的和本文的 GeBP 转录因子蛋白都有 DUF573 功能结构域(陈凯等，2019；单雪萌等，2020；图 6)，尽管该结构域的功能未知，推测该功能域应该位于转录因子的中央 DNA 结合区，可以与靶基因的顺式作用元件结合调控基因的表达，但是因为 DUF573 功能结构域在转录因子上的位置不同，所以在调控靶基因的表达上也存在差异。

基因的表达模式同基因的功能是紧密相连的。总体来看，不同物种 *GeBP* 转录因子基因在各自物种不同组织和不同发育时期都有表达，很多只是表达强弱的差异(陈凯等，2019；单雪萌等，2020；图 7)，表明 *GeBP* 转录因子基因在不同组织和不同发育时期都发挥有重要的功能。对毛竹 16 个 *PeGeBPs* 基因的表达研究发现，其中有 12 个 *PeGeBPs* 基因在带有表皮毛的叶、箨、箨片以及纤毛中的表达量高于无表皮毛的笋，表明它们在表皮毛的形成中应该发挥了主要功能(单雪萌等，2020)。表皮毛在大豆植株上的分布位于叶片、茎秆、豆荚和花萼等地上部器官表面，结合 *GeBP* 转录因子基因调控表皮毛生长发育的功能，*GmGeBP7* 和 *GmGeBP6* 是研究大豆调控表皮毛生长发育的最佳候选基因。

本研究通过生物信息学的方法，从植物转录因子数据库和大豆基因组搜寻获得 9 个 *GeBP* 转录因子基因，随后对家族成员基因的核苷酸序列长度和氨基酸序列的基本理化性质、基因在染色体的物理定位、基因结构、系统进化树、序列对比、功能结构域分布、基因在不同组织的表达模式进行综合的预测和分析，研究结果将为进一步深入探究大豆 *GeBP* 转录因

子基因的功能机制提供理论依据和参考价值。

参考文献：

CARLES CC, CHOFFNES-INADA D, REVILLE K, et al., 2008. ULTRAPETALA1 encodes a SAND domain putative transcriptional regulator that controls shoot and floral meristem activity in Arabidopsis[J]. Development, 132(5): 897-911.

CHEN K, LIU JQ, SONG HH, et al., 2017. Identification, evolution and expression analysis of GeBP transcription factors family in tomato[J]. Mol Plant Breed, 15(9): 3438-3445. [陈凯，刘金秋，宋海慧，等，2017. 番茄 GeBP 转录因子家族的鉴定及其进化和表达分析[J]. 分子植物育种，15(9)：3438-3445.]

CHEVALIER F, PERAZZA D, LAPORTE F, et al., 2008. GeBP and GeBP-like proteins are noncanonical leucine-zipper transcription factors that regulate cytokinin response in arabidopsis[J]. Plant Physiol, 146(3): 1142-1154.

CURABA J, HERZOG M, VACHON G, et al., 2003. GeBP, the first member of a new gene family in Arabidopsis, encodes a nuclear protein with DNA-binding activity and is regulated by KNAT1[J]. Plant J, 33(2): 305-317.

CURABA J, MORITZ T, BLERVAQUE R, et al., 2004. *AtGA3ox2*, a key gene responsible for bioactive gibberellin biosynthesis, is regulated during embryogenesis by LEAFY COTYLEDON2 and FUSCA3 in Arabidopsis[J]. Plant Physiol, 136(3): 3660-3669.

HU B, JIN J, GUO AY, et al., 2015. GSDS 2.0: an upgraded gene feature visualization server[J], Bioinformatics, 31(8): 1296-1297.

JASINSKI S, PIAZZA P, CRAFT J, et al., 2005. KNOX action in Arabidopsis is mediated by coordinate regulation of cytokinin and gibberellin activities[J]. Curr Biol, 15(17): 1560-1565.

JIANG J, MA S, YE N, et al., 2017. WRKY transcription factors in plant responses to stresses[J]. J Integr Plant Biol, 59(2): 86-101.

KUMAR S, STECHER G, LI M, et al., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms[J]. Mol Bio Evol, 35(6): 1547-1549.

LIU B,QIU S,HE JQ, et al., 2020. Bioinformatics analysis and expression of eight Dof transcription factors in soybean under drought stress[J]. Soybean Sci, 39(3): 377-383. [刘蓓，邱爽， 何佳琦，等，2020. 8 个大豆 Dof 转录因子的生物信息学分析及干旱诱导表达[J]. 大豆科学，39(3)：377-383.]

NOH B, LEE SH, KIM HJ, et al., 2004. Divergent roles of a pair of homologous jumonji/zinc-finger-class transcription factor proteins in the regulation of Arabidopsis flowering time[J]. Plant Cell, 16(10): 2601-2613.

225-231.

PU L, SUO JF, XUE YB, 2003. Molecular control of plant trichome development[J]. Acta Genet Sin, 30(11): 1078-1084. [普莉，索金凤，薛勇彪，2003. 植物表皮毛发育的分子遗传控制[J]. 遗传学报，30 (11)：1078-1084.]

RAY S, DANSANA PK, GIRI J, et al., 2011. Modulation of transcription factor and metabolic pathway genes in response to water-deficit stress in rice[J]. Funct Integr Genomics, 11(1): 157-178.

SARKER LS, ADAL AM, MAHMOUD SS, 2019. Diverse transcription factors control monoterpene synthase expression in lavender (Lavandula)[J]. Planta, 251(1): 1-5.

SHAN XM, YANG KB, SHI JJ, et al., 2020.Genome- wide identification and expression analysis of GeBP transcription factor gene family in moso bamboo[J]. J Nanjing For Univ Nat Sci Edi,

44(3): 41-48. [单雪萌，杨克彬，史晶晶，等，2020. 毛竹 GeBP 转录因子家族的全基因组鉴定和表达分析[J]. 南京林业大学学报(自然科学版)，44(3)：41-48.]

SHI L, 2013. Preliminary functional analysis of the GeBP gene family in rice[D]. Wuhan: Huazhong Agricultural University. [石蕾，2013. 水稻 GeBP 家族基因的功能初探[D]. 武汉：华中农业大学.]

SONG H, WANG PF, HOU L, et al., 2016. Global Analysis of WRKY genes and their response to dehydration and salt stress in soybean[J]. Front Plant Sci, 7: 9.

KENICHI T, TOSHIHIRO T, SUSUMU H, et al., 2004. Three Arabidopsis MBF1 homologs with distinct expression profiles play roles as transcriptional co-activators[J]. Plant Cell Physiol, 45(2): 225-231

YU YC, WANG N, HU RB, et al., 2016. Genome-wide identification of soybean WRKY transcription factors in response to salt stress[J]. Springerplus, 5(1): 920.

ZHAI Y, QIU S, ZHANG J, et al., 2019. Bioinformatics and expression analysis of three Dof transcription factors in soybean[J]. Acta Agric Boreal-Sin, 34(6): 14-19. [翟莹，邱爽，张军，等，2019. 大豆中 3 个 Dof 转录因子的生物信息学及表达分析[J]. 华北农学报，34(6)：14-19.]

ZHAI Y, ZHANG J, ZHAO Y, et al., 2016. Bioinformatics and expression analysis of 5 newfound ERF genes in soybean[J]. Acta Agric Zhejiangensis, 28(10): 1644-1649. [翟莹，张军，赵艳，等，2016. 大豆 5 个新发现 ERF 基因的生物信息学及表达分析[J]. 浙江农业学报，28(10)：1644-1649.]